ETL-0516

AD-A204 167

Dynamic Image Interpretation for Autonomous Vehicle Navigation 1987 End of Year Technical Report

Edward M. Riseman
Allen R. Hanson

University of Massachusetts
Computer & Information Science Department
Amherst, Massachusetts 01003

DTIC
ELECTE
JAN 24 1989
S D
H

September 1988

Approved for public release; distribution is unlimited.

Prepared for:

Defense Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, Virginia 22209-2308

U.S. Army Corps of Engineers
Engineer Topographic Laboratories
Fort Belvoir, Virginia 22060-5546

89 1 19 040

Destroy this report when no longer needed.
Do not return it to the originator.

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

The citation in this report of trade names of commercially available products does not constitute official endorsement or approval of the use of such products.

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION
UNCLASSIFIED

1b. RESTRICTIVE MARKINGS

2a. SECURITY CLASSIFICATION AUTHORITY

2b. DECLASSIFICATION / DOWNGRADING SCHEDULE

3. DISTRIBUTION / AVAILABILITY OF REPORT
Approved for public release;
distribution is unlimited.

4. PERFORMING ORGANIZATION REPORT NUMBER(S)

5. MONITORING ORGANIZATION REPORT NUMBER(S)
ETL-0516

6a. NAME OF PERFORMING ORGANIZATION
University of Massachusetts

6b. OFFICE SYMBOL *(If applicable)*

7a. NAME OF MONITORING ORGANIZATION
U.S. Army Engineer Topographic Laboratories

6c. ADDRESS *(City, State, and ZIP Code)*
Computer & Information Science Department
Amherst, Massachusetts 01003

7b. ADDRESS *(City, State, and ZIP Code)*
Fort Belvoir, VA 22060-5546

8a. NAME OF FUNDING / SPONSORING ORGANIZATION
Defense Advanced Research Projects Agency

8b. OFFICE SYMBOL *(If applicable)*

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
DACA76-85-C-0008

8c. ADDRESS *(City, State, and ZIP Code)*
1400 Wilson Boulevard
Arlington, VA 22209-2308

10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO
62301E			

11. TITLE *(Include Security Classification)*
Dynamic Image Interpretation for Autonomous Vehicle Navigation
1987 End of Year Technical Report

12. PERSONAL AUTHOR(S)
Edward M. Riseman & Allen R. Hanson

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Annual	FROM 2/26/87 TO 2/25/88	1988 September	32

16. SUPPLEMENTARY NOTATION

17. COSATI CODES

FIELD	GROUP	SUB-GROUP

18. SUBJECT TERMS *(Continue on reverse if necessary and identify by block number)*
Scene Interpretation
Sensor Motion
Spatial Reasoning

19. ABSTRACT *(Continue on reverse if necessary and identify by block number)*

This report presents the results of the project on Dynamic Image Interpretation for Autonomous Land Vehicle (ALV) Navigation for the time period 2/26/87--2/25/88. The purpose of the ALV project is to develop algorithms and tools to enable a vehicle to navigate autonomously through realistic landscapes.

The current effort continues our work on this problem. In the report we summarize the accomplishments of the past year in constructing robust algorithms to be used for vehicle navigation, as well as tools that have been developed to more efficiently utilize these algorithms.

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT
☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS

21. ABSTRACT SECURITY CLASSIFICATION
UNCLASSIFIED

22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE (Include Area Code)	22c. OFFICE SYMBOL
Linda H. Graff	202-355-2700	CEETL-RI

DD Form 1473, JUN 86 *Previous editions are obsolete.*

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

End Of Year Technical Report:
Dynamic Image Interpretation For
Autonomous Vehicle Navigation

Contract: DACA76-85-C-0008
February 26, 1987 - February 25, 1988

Principal Investigators:

Edward M. Riseman and
Allen R. Hanson

Computer and Information Science Department
University of Massachusetts at Amherst
Amherst, MA 01003

Contents

1 Visual Motion Analysis **3**

1.1 Computation of the Optical Flow Field 3

1.2 The Recovery of Environmental Motion and Structure from a Mobile Vehicle 4

1.3 Alternatives to General Motion Analysis 6

1.3.1 Stereoscopic Motion Analysis 6

1.3.2 Analysis of Constant General Motion 7

1.4 Token-Based Approaches to Motion and Perceptual Organization 7

2 Mobile Vehicle Navigation **9**

3 Perceptual Organization (Grouping) **11**

3.1 The Perceptual Organization of Image Curves 11

3.2 Extracting Geometric Structure 11

4 Database Support for Symbolic Vision Processing **12**

4.1 ISR1 14

4.2 ISR2 15

4.3 Generic Views and Indexing 16

Background

Since the beginning of this contract, our research group has developed a variety of motion algorithms, and in most cases applied them to real-world image sequences, including domains of robot arm workspaces, indoor hallways, and outdoor sidewalk/road scenes. In particular, experimental investigations of translational motion sequences demonstrated some degree of robustness. Anandan [ANA85b], [ANA87a], [ANA87b], [ANA88] developed an algorithm for determining feature point correspondences between frames that allowed the computation of dense displacement fields with associated confidences. This capability could be used to effectively track points across frames. Lawton [LAW83], [LAW84] showed that the focus-of-expansion (FOE) often could be extracted from a sensor undergoing pure translational motion (i.e two degrees of freedom) to within a few degrees of accuracy. Glazer [GLA87a], [GLA87b], in his recently completed Ph.D. thesis, developed two algorithms for the efficient computation of image motions using hierarchical multiresolution methods operating over image data pyramids.

Bharwani [BHA85], [BHA86] developed a multi-frame algorithm for depth extraction under known translational motion which iteratively predicts the image motion of a feature point in future frames, determines correspondence by a search over the limited predicted area, and then refines the depth estimate using the new match. Snyder [SNY86] analyzed the effects of uncertainty in the location of the FOE and feature points in the image on the computation of depth, and showed how this analysis could be used to quantitatively provide predictions for constraining the search window used for matching these points in future frames.

Adiv [ADI85a], [ADI85b], [ADI85c] developed an algorithm for general sensor motion (five degrees of freedom) in an environment with objects undergoing independent general motion, the goal being to recover the motion parameters of both the sensor and any visible moving objects. This latter problem is much harder, and although there was some empirical demonstration of capabilities, there was an assumption that this algorithm would be computationally more complex, and perhaps less robust, than the algorithms for translational motion.

Over the past several years, we have developed the notion of a 'schema' as the basic unit of knowledge representation in the VISIONS system. Within the schema system image interpretation is the process of instantiating a subset of schemas to build a description of the three-dimensional scene which gave rise to the image. Knowledge is represented in a 3 level abstraction hierarchy of schema nodes by part/subpart descriptions, class/subclass descriptions, and expected relationships between schemas; the resultant hierarchical graph constitutes the VISIONS knowledge network. [HAN78b], [HAN78c], [HAN87a], [DRA87b].

The VISIONS system is organized around three levels of data representation and types of processing. At the low-level, the representations are in the form of numerical arrays of sensory data with processes for extracting the image events that will form the intermediate representation. At the intermediate level, the representation is composed of symbolic tokens representing regions, lines, surfaces and the attributes of these primitive elements (which might include local motion and depth information). The intermediate representation is stored in a data base called the intermediate symbolic representation (ISR) which supports grouping (perceptual organization) and information fusion processes that are employed to develop aggregations of existing tokens to form new tokens. At the high level, the representation is a set of object hypotheses and active schema instances which control the intermediate and low-level processes. Control initially proceeds in a data-directed manner and later is significantly top-down in a knowledge-directed manner.

Based on our experience with an initial implementation of the schema system and a set of experiments designed to interpret reasonably complex house scenes [HAN86], [HAN87a], [WEY86] a new schema system and support environment has been designed and partially implemented [DRA87a]. Two new tools, the Intermediate Symbolic Representation and the Schema Shell, have been developed and are currently being tested and extended using the interpretation of road scenes as a second experimental task domain.

We view the task of perceptual organization and grouping as the extraction of relevant structure from overfragmented and incomplete descriptions and the construction of more abstract descriptions from less abstract ones. By this we mean algorithms which have as input the tokens produced by the low-level system and other grouping operations (region, lines, flow fields,...) and have as output more complex tokens generated by grouping strate-

gies based on the *relations* between the tokens. The goal of this type of 'intermediate' level processing is the reduction of the substantial representational gap which exists between the low level image descriptors and the primitives with which the high level semantic descriptions are constructed. The process of abstraction thus involves the search for events which can be more concisely described as a unit and which results in a description which may be more relevant to the evolving semantic interpretation.

This is the approach taken by Boldt and Weiss [WEI86] who developed a scale-sensitive hierarchical algorithm for grouping collinear line segments into progressively longer segments on the basis of geometric properties of the hypothesized group as well as the similarity of image features along both sides of the component lines.

1 Visual Motion Analysis

We have continued our analysis of motion, with an emphasis on techniques which will be of practical use for autonomous navigation. Our effort has been concentrated in several directions which build on previous work done at the University of Massachusetts: the computation of the optical flow field and the design of practical, robust algorithms for determining the structure of the environment from a mobile vehicle.

1.1 Computation of the Optical Flow Field

In his recently completed doctoral dissertation [ANA87a] Anandan provides a unified framework for extracting a dense optical flow field from a pair of images, as well as an integrated system which is based on a matching approach (see, also, [ANA85a], [ANA85b] [ANA87a], [ANA87b]). This framework appears to be sufficiently general to encompass both gradient-based and correlation matching approaches. It consists of a hierarchical scale-based matching scheme using bandpass filters, orientation-dependent confidence measures, and a smoothness constraint for propagating reliable displacements. His integrated system [ANA85a] for the extraction of displacement fields uses the minimization of the sum-of-squared-differences (SSD) measure as the local match criterion, and computes confidence measures based on the shape of the SSD surface. It also formulates the smoothness

odes
/or
Dist Special
A-1
DTIC COPY INSPECTED 4

assumption as the minimization of an error functional and overcomes many of the difficult problems that exist in other techniques. The error functional consists of two terms. One, called the *approximation* error, measures how well a given displacement field approximates the local match estimates, while the other, called the *smoothness* error, measures the global spatial variation of a given displacement field. The finite-element method is used to solve the minimization problem. The approach also gives information for extracting occlusion boundaries in some situations.

Anandan has also shown [ANA84] that the functional minimization problem formulated in his matching technique converges to the minimization problem used in gradient-based techniques (e.g. Glazer's technique discussed later). In particular, by relating an approximation error functional used in his matching approach to the intensity constraints used in the gradient-based approaches, he explicitly identifies confidence measures which have thus far been implicitly used in the gradient-based approach. Finally, he suggests ways that algorithms operating on a pair of frames can be developed into multiple-frame algorithms and discusses their relationship to spatio-temporal energy models.

Glazer's recently completed thesis [GLA87c] presents an approach to motion detection using multi-resolution methods in a hierarchical processing architecture. Two motion detection algorithms are developed and analyzed. The hierarchical correlation algorithm utilizes a coarse-to-fine control strategy across the resolution levels and overcomes two disadvantages of single-level correlation: large search areas, which require expensive searches, and repetitive image structures, which cause incorrect matches. The hierarchical gradient-based algorithm [GLA87a],[GLA87b], generated over low-pass image pyramids, extends single-level gradient algorithms to the computation of large displacements. Within each level the next refinement of the displacement field is obtained by combining a local intensity constraint and a global smoothness constraint. The mathematical formulation involves the minimization of an error functional consisting of two terms, corresponding to the intensity and the smoothness constraints mentioned above. The minimization problem is solved using the finite-difference approach, which leads to a multi-resolution relaxation algorithm. A formal analysis of the hierarchical gradient algorithm is presented, including the basic equations for computing a refined disparity vector, the discrete representations

and computations for solving these equations, and a geometric interpretation of the resulting relaxation algorithm. The experimental results show that the two algorithms have comparable accuracy and a cost analysis shows that the hierarchical gradient algorithm is less costly.

1.2 The Recovery of Environmental Motion and Structure from a Mobile Vehicle

Our previous research in motion analysis led us to attempt to deal with a real application subsystem for the CMU NAVLAB [THOM87]. The goal was to detect obstacles in the path of the vehicle at distances beyond the limits of the ERIM range sensor, i.e. at distances beyond 40 feet. Initial results from Bharwani's algorithm [BH 85], [BHA86] implied the possibility of extracting usable depth of obstacles at distances between 40 and 80 feet. By applying an FOE extraction algorithm prior to the depth extraction algorithm, there was the expectation that an effective subsystem could be developed. To accomplish this in actual imaging situations on a moving vehicle turned out to be far more difficult than we expected.

In dynamic imaging situations where the sensor is undergoing primarily translational motion with a relatively small rotational component (i.e. "approximate" translational motion), it might seem likely that translational motion algorithms would be effective in determining depth. Although translational motion is the dominant form of motion and is approximately constant over a long sequence of frames, there usually are local variations due to irregularities in the road surface (bumps, holes, and undulations), as well as minor rotation of the vehicle as it translates. This is often manifested by changes in the location of the FOE (i.e. effectively it produces a different translational motion), and in rotational motions that must be removed if correct values of depth are to be extracted from the feature displacements. An attempt to correct for these effects via a relatively simple preprocessing "registration" algorithm [SNY86] without utilizing full analysis of the general motion problem also led to difficulties, even when the rotations were as small as 0.1° to 0.5°. This registration algorithm consisted of two parts. In the first part, the motion of distant points was used to find the rotational component of the vehicle's motion, and in

the second part, this rotational component was subtracted from the full optical flow field, leaving a flow field that is (in principle) due to purely translational motion.

The basic problem with this registration algorithm was that the rotational parameters of the vehicle's motion could not be found accurately, and hence subtraction of the rotational component of the motion from the optical flow field resulted in a flow field containing a residual rotational component. Our experimentation with the registration algorithm as a preprocessing step to the Bharwani algorithm led us to conclude that the Bharwani algorithm (as well as any other algorithm assuming purely translational motion) was very sensitive to this residual rotation. We discuss these issues in greater detail in the paper of Dutta, et al. [DUT88]. We also explain in this paper the theoretical reasons why it is unlikely that algorithms based on purely translational motion will work except in very particular and restricted environments.

The theoretical and experimental problems we encountered with using purely translational motion algorithms led us to combine the optical flow algorithm of Anandan [ANA85a] and the algorithm of Adiv [ADI85a] to obtain an algorithm for computing environmental motion and structure for the case where both translations and rotations are present, i.e., a general motion algorithm. We applied this algorithm to the same dynamic image sequences for which the combination of the registration and Bharwani algorithms failed to recover environmental motion and structure accurately, and found that results were significantly improved by using the general motion algorithm. In particular, the general motion algorithm recovered the depths of objects in an image sequence taken from the CMU NAVLAB to within an average error of 10% (see [DUT88] for further details).

The conclusion we draw from this analysis is that in many real situations general motion analysis must be applied in order to determine depth of points, even when sensor motion is primarily translational with only small amounts of rotation. One obvious hardware solution (at significantly increased cost) is the use of a gyro-stabilized platform or a land navigation system to recover translational and rotational motion so that sensor motion typically will be much closer to the case of pure translational motion. In the next section we discuss alternative approaches for the extraction of motion parameters and depth. We will be exploring the utility of these and the general motion algorithm discussed above in

the continuation of our work on the Autonomous Land Vehicle.

1.3 Alternatives to General Motion Analysis

1.3.1 Stereoscopic Motion Analysis

By carrying out motion analysis with a pair of cameras - stereoscopic motion - the additional constraints can significantly reduce the complexity of the analysis on a theoretical level. Balasubramanyam and Snyder [BAL87a], [BAL87b] have developed an algorithm to extract the parameters of *motion in depth*: the single component of translation in depth (i.e. parallel to the line of sight) and the two components of rotation in depth (i.e. rotations that are not around the line of sight). This is achieved by building upon the work of Adiv for segmenting the flow field into rigid independently moving objects [ADI85a], and the formulation of Waxman and Duncan [WAX86], which shows that the ratio of the relative optic flow between a stereo pair of images to the disparity between them is a linear function of the image coordinates. Experimental results are presented for simulated data of general motion of both the sensor and independently moving objects. Work is currently underway to test the effectiveness of this algorithm on real scenes.

1.3.2 Analysis of Constant General Motion

Another way to introduce additional constraints to the problem of general motion analysis in an effort to achieve practical, robust algorithms is via Shariat's formulation: constant but arbitrary general motion of a rigid object [SHA86]. This leads to a set of difference equations across a sequence of images, relating the positions of a feature in the image plane to the motion parameters of the projected point. The solution obtained is a set of 5th order non-linear polynomial equations in the unknown motion parameters, whose solution requires a Gauss-Newton non-linear least-squares method with carefully designed initial guess schemes. Pavlin [PAV85] has derived a closed-form solution for the rigid object trajectory by integrating the differential equations describing the motion of a point on the tracked object. The integrated equations are non-linear only in angular velocity, and are linear in all other motion parameters. These equations allow the use of a simple

least-square error minimization criterion in an iterative search for the motion parameters.

1.4 Token-Based Approaches to Motion and Perceptual Organization

The problems cited in Section 1.2 with respect to the extraction of motion and depth information using traditional optical flow techniques have led us toward the exploration of methods for combining the local flow/displacement fields with larger token-like structures. It is our position that the inherently local measurement of visual motion provided by optical flow is insufficient to meet the varied requirements of dynamic image understanding. The approach we are developing involves computing the correspondence between tokens of arbitrary spatial scale produced by perceptual organization processes. Such tokens often map directly to environmental structure, and descriptions of their movement often correlate more closely with the motion of physical objects than does the local motion information contained in the displacement field. A token match represents more than just a spatial displacement: also explicit in this representation are the time-varying values of those parameters which define the token or which can be extracted from the structure of the token.

Williams and Hanson [WIL88a],[WIL88b] describe work in progress toward this goal. The premise of this work is that the structure obtained from perceptual organization processes can be combined with the local motion information contained in the flow field to provide a more robust estimate of motion and depth parameters. The approach can be viewed as augmenting the rather limited use of spatial structure in traditional approaches with the richer descriptive vocabulary of spatial structure provided by the perceptual organizational processes over both space and time. In this sense, the spatially organized structures (such as lines, regions, curves, vertices, intersections, rectangular groups, etc.), which are actively constructed from the image, can be considered to be interest operators of large spatial extent.

In the first pap[er] [WIL88a], a method for computing the temporal correspondence between straight line segments is presented. We consider the two frame case here, but the method is extensible and has been extended to multiple frames. A straight line perceptual

organization process, developed by Boldt and Weiss [BOL87], [WEI86], is applied to both frames independently to provide straight lines in each frame. A displacement field is also computed from the two frames using the algorithm developed by Anandan [ANA87b], [ANA88] described above. After filtering the straight lines on length and contrast to reduce the line set in both images, the displacement field is used to construct a search area in frame 2 for each line in frame 1. Since a one-to-one correspondence between lines is unlikely, a minimal mapping approach [ULL79] is used to compute the correspondence between the frame 1 and frame 2 line sets; such a mapping is called a minimal bipartite cover. The similarity measure used to compute the cover involves the similarity and spatial separation of the candidate token matches. By computing the connected components of the bipartite graph, the global matching problem is conveniently divided into smaller, individually tractable pieces which reflect the scope of potential interactions; a simple blind search of the subgraphs is used to extract the bipartite cover minimizing the positional and similarity discrepancy metric.

The matching results obtained are quite good. The system has been run repeatedly on successive frames of several multi-frame sequences. In the multi-frame case, a directed acyclic graph is constructed which represents the splitting and merging patterns of line segments over time. Work is in progress to analyze the trajectories of the tokens over time.

In the second paper [WIL88b], a method for computing depth from the line correspondences is described using the temporal change in the length of virtual lines constructed from the intersections of the Boldt lines [BOL87]. We use virtual lines because the length of the original lines is not reliable, although their orientation and lateral displacement is quite precise. This "looming" method is also generalized to areas. The method is generally applicable to structures whose total extent in depth is small compared to the depth of its centroid (that is, for those cases in which perspective projection can be approximated by scaled orthographic projection [THOM87]) and which do not exhibit any independent motion. The technique does not depend on the complete determination of egomotion parameters of the sensor, but it does require the computation of the translation component of the sensor in the direction of motion. An analysis of the sensitivity of the algorithm to errors in the measured variables is planned for the near future; experimental results on

real image sequences have shown that the algorithm may be quite robust.

2 Mobile Vehicle Navigation

The hardware platform for experimentation in mobile robotics at UMass is a Denning Mobile Robotics vehicle with a B&W television camera and UHF transmitters and receivers for uplink and downlink communication to a Gould IP8500 image processing system connected to a Vax 11/750 computer. Plans are underway to utilize a 12-node Sequent multiprocessor to improve the computational effectiveness of our experimentation environment.

Arkin [ARK87a, ARK87b, ARK87c] used this platform to develop AuRA (Autonomous Robot Architecture), which integrates planning, cartographic, perception, motor, and homeostatic systems into a functional robot navigation system. The system is designed to navigate in the hallways and outdoor environment surrounding our building at UMass.

Aura employs a 'meadow' map as its long-term memory; the meadow map is used for global path planning and contains embedded a priori knowledge to guide sensor expectations used for positional updating. A layered short-term memory based on instantiated meadows represents the currently perceived world. A hierarchical path planner produces a global path free of collisions with all modeled obstacles.

Aura extends the idea of schemas, as currently employed in the VISIONS system, to include the mobile robot domain. The schema-based path execution system handles unexpected and dynamic obstacles not present in the robot's world model. This motor-schema based navigation system produces reactive/reflexive behavior in direct response to sensor events. In addition, new techniques in the treatment of robot uncertainty, which expedite sensory processing, were developed. These include the use of a spatial error map with associated growth and reduction techniques.

Several computer vision sensor strategies have been developed for use within Aura. These include a fast line finding algorithm that is a simplified and more efficient version of the Burns straight line extraction algorithm (at the price of robustness) [BUR86, KAH87].

a fast simplified region segmentation algorithm based on the VISIONS region segmentation system [BEV87], and a depth from motion algorithm [BHA86]. Aura uses both vision and ultrasonic sensing during path traversal.

We are currently rebuilding Aura to make better use of the information available from the visual sensors and to more completely integrate the full spectrum of image understanding techniques developed in the VISIONS project. In particular, we intend to utilize some of the depth from motion algorithms discussed above [BAL88], [DUT88], [WIL88a], [WIL88b]; and some of the simpler object recognition strategies of the schema system [DRA87a], [DRA87b], [HAN87b], including strategies for multi-sensor information fusion [BEL86], [RIS87].

3 Perceptual Organization (Grouping)

3.1 The Perceptual Organization of Image Curves

Most of our work in perceptual organization [BEL86], [BOL87],[BUR86],[DOL86],[DOL88], [REY84], [REY86b],[REY87],[RIS87],[WEI85],[WEI86], [WIL88a],[WIL88b] has been focussed on rectilinear structures (e.g. straight lines, corners, parallel line pairs, and the like). Of course not all of the world can be described by straight lines. Consequently, Dolan [DOL88] has been exploring methods for extending the general technique developed by Boldt [BOL87],[WEI86] to the simultaneous extraction of curves, straight lines, and corners (including cusps); these are the primitive descriptive elements. The basic operation cycle consists of linking, grouping, and replacement, which takes place at increasing perceptual scales, resulting in a hierarchical scale-space description of these important image events.

The linking stage finds subsets within the set of initial local edge tokens that satisfy the binary constraints of the particular grouping principles employed. The grouping mechanisms perform a detailed geometric analysis on sets of linked tokens whose extent is within the current scale; in Dolan's system, this also entails classification and ranking of the token sequences as one of the basic primitive elements. Replacement mechanisms encode the geometry of a surviving group by substituting a single token for the group. The process

then repeats at the next scale.

3.2 Extracting Geometric Structure

Reynolds and Beveridge [REY87] have been developing a perceptual grouping system for the extraction of rectilinear structures from an initial set of line primitives obtained using the straight line extraction algorithm developed by Burns, Hanson, and Riseman [BUR86]. The lines are represented as nodes in a graph. The grouping criteria are the geometric relations of spatially proximate collinear, spatially proximate parallel, spatially proximate orthogonal, or any subset of these relations; the relations form the links in the graph.

Line groups are generated using a connected components analysis of the chosen geometric links. Finally, individual geometric structures (e.g. rectangles, collinear lines, parallel line pairs, corners, etc.) may be identified as subgraphs of the connected components. These techniques have been applied to extraction of objects such as road networks in aerial images.

Object recognition strategies can be represented as relational graphs to be matched to extracted data. The problems associated with fragmentation, as well as merged and missing tokens, makes this a difficult problem. However, multiple representations (such as lines and regions) can be brought together to provide partial redundancy [RIS87]. Thus, current work overlaps issues of constrained graph matching, perceptual organization, and information fusion.

4 Database Support for Symbolic Vision Processing

It is becoming increasingly evident that intermediate-level vision, and the perceptual grouping processes encompassed, are an extremely important component of any knowledge-based interpretation system. Our current view is that a major goal of the perceptual organization processes is to reduce the substantial gap which exists between the extracted image descriptors and the high level knowledge representations of the objects. The more abstract the intermediate level tokens are, the more computationally efficient the matching

is between high level descriptions and the intermediate level tokens, where general world knowledge is used to constrain the set of possible interpretations.

The intermediate level may be viewed as simply a symbolic representation of primitive image 'events' as points, regions, lines, contours, areas, surfaces, etc. and their features, created by an iconic to symbolic transformation of the image data. However, recent work in vision has shown it to be much more than a passive level of data representation. Many of the recently developed grouping operators, for example, function at the intermediate level by building more abstract structures from the primitive descriptions [BOL87],[DOL88], [DOL86], [FIS86], [LOW85], [REY87], [WIL88b], [WIT83].

Consequently, we view the intermediate level as hosting active processes which construct more abstract tokens from less abstract ones. Universally applicable similarity operators and geometric constraints are employed on the evolving spatial structures. In order to facilitate research on image interpretation systems, where data and control are closely coupled throughout all three stages, mechanisms must be provided for efficient structuring of the data and processes.

In addition, the complexity of many vision systems requires the cooperation and interaction of many researchers and the integration of their subsystems. The applications are far too large for an individual to solve on his own. Thus, the intermediate level representations and software environment must support, at a minimum, the following:

- a single uniform data interface to both high and low levels;
- sharing of data between levels, and between researchers at all levels;
- integration of research results into a monolithic system;
- standard handling of common relational and geometric queries, to reduce the programming overhead of coding them from scratch;
- distribution of data and processes over several machines and in several computer languages (C, LISP, FORTRAN)
- an efficient programming environment for intermediate level algorithm development.

Unfortunately, current understanding of this level of vision makes it impossible to predict the kind of structures which must be represented, the types of access to these

structures, the kinds of relationships which might exist between them, or the range and type of descriptive features attached to them. At this point it appears that quite a diverse set of representations and mechanisms are employed in various vision system components. We can minimally assume that the intermediate level must support known methods of information fusion and perceptual organization, and provide the flexibility to support the representation and manipulation of geometric and structural relations. For example, the types of data which should be representable at this level include:

- points: endpoints, points of high curvature, vertices, virtual points, etc;
- lines and curves: edges, straight lines, curve segments;
- areas: regions, surface patches, focus of attention areas, etc.;
- relations: adjacency, containment, intersection, etc;
- structures: grouped lines and edges, edge-vertex tuples (e.g.corners), line chains, geometric structures, and generally subsets of tokens defined by a relation.

Each has an associated set of features, or descriptors, whose definition may vary as research progresses. Consequently, there are two fundamental types of data access that must be supported: access to tokens by name and by feature value (associative access); note that we also treat relations as features. It is rarely the case that a token definition stays constant over the course of an interpretation. Tokens may be split from or merged with other tokens, features recomputed, and tokens may take part in many set relationships with other tokens.

4.1 ISR1

Research into intermediate level grouping mechanisms [BOL87], [DOL88], [FIS86], [LOW85], [REY87], [WIL88b], [WIT83] and the development of the VISIONS schema system [DRA87a], [DRA87b], [DRA88], [HAN78a], [HAN78b], [HAN78c], [HAN86] have led us toward the development of a flexible and efficient intermediate level of representation called the Intermediate Symbolic Representation (ISR) [BRO88], [DRA87a], [HAN87a].

ISR1 was implemented in 1985 primarily as a data interface between the output of the low-level image segmentation and feature extraction processes running in C on a DEC

VAX and the high-level symbolic interpretation system running in Lisp on a TI Explorer. The unit of representation in ISR1 is the token, composed of a name and a list of features. The features are described through a lexicon and tokens sharing a common lexicon are organized into a tokenset. Each feature entry in the lexicon consists of a datatype and an optional on-demand function for computing the feature value. Standard feature datatypes include type real, integer, pointer, extents, and bitplane. Extents is simply the coordinates of the bounding rectangle of the token in the image plane. Features of type bitplane are binary masks defining the spatial coverage of the token in the image.

Since a tokenset may be viewed as a two-dimensional array, access to elements in the array are by token name (the rows) and constraints on feature values (the columns). Associative access of elements are returned as a list or as an array. One of the major design deficiencies of ISR1 was that there were no convenient mechanisms for representing and storing these lists of elements.

4.2 ISR2

ISR1 was used heavily over a period of years by researchers whose individual research focus was distributed reasonably uniformly over all three levels of abstraction. During this period of time, a number of design deficiencies were noted in ISR1; two of the major problems which necessitated the redesign were:

- The separation of the lexicon from a tokenset created problems. When the lexicon had to be modified, old tokensets no longer had valid descriptions; (for example, when a feature was added to a set of region tokens). Short-term solutions resulted in a proliferation of stored tokensets and a great deal of confusion at the application level.
- Sets of associatively accessed tokens could not be conveniently manipulated, made into tokensets, nor stored as features of other tokens. In particular, it was difficult to relate tokens across token types (such as regions and lines).

In response to these problems a decision was made to design a new version of the symbolic database [BRO88]. ISR2 retains the basic flavor of ISR1, including tokensets, the basic token access functions, and features and feature datatypes. The lexicon concept was eliminated in favor of associating the feature descriptions with the tokenset itself.

Recognizing that there were other pieces of information which apply to the tokenset as a whole (such as generation dates, image information, and processing history), each tokenset is now organized as a simple frame, with slots for the various features of the tokenset. Frame features include the simple types integer, real, string, frame, and tokenset and the complex types composite, sort, slice, and virtual. The frame feature allows frame hierarchies to be constructed. The tokenset feature points to the tokenset or tokensubset associated with the frame. The composite feature is a generalization of feature types like bitplane and extents from ISR1 (i.e. they are multi-valued features). Virtual features are features whose values can be calculated but not stored, hence they are always calculated on demand. They serve much the same purpose as methods in an object-oriented programming language. Sort and slice datatypes provide facilities for defining and maintaining partitions based on feature values; for example, a typical application for a slice feature might be to create and maintain a grid for fast 2D spatial access to tokens from the image coordinates. Other modifications to ISR1 include a more comprehensive file management system to deal with the frame hierarchies, the addition of several types of demons (on-demand functions), and extensions to the command language to support the new capabilities.

Like ISR1, much of ISR2 will be implemented in C with a LISP user interface. Implementation is now underway and testing will begin in the near future. Since vision is such a dynamic research environment, it would not be unreasonable to expect ISR3 after experience with ISR2 is obtained.

4.3 Generic Views and Indexing

Given a large number of models it is not possible to match each model with the data. Regardless of the type of model being used, it becomes necessary to use image features or tokens as an index into the model base. We have formulated a solution to this problem in terms of generic views (also known as characteristic views or aspect graphs) [CAL85], [CHA82], [FEK84], [IKE87], [KER81], [KOE84], [KOE76], [KOE79], [STE87]. Generic views can be seen as an intermediate representation providing the link between a solid modeling system and a predictive graph structure to be used in matching. A generic view representation is a set of typical views of an object or parts of an object which allow one to

distinguish that object from others in a given model base and from any viewing direction which does not involve some special alignment.

The method of generic views is defined here for a system in which images are obtained by orthogonal projection, but it has been generalized to perspective projection. Imagine a sphere centered on and enclosing a 3D object model. Each point on this *viewing sphere* corresponds to a unique view of the object, and hence to a unique parameterized 3D to 2D projection. Given a set of features (which can include relational features), one can *a priori* divide up the viewing sphere into *generic sectors* by grouping those directions for which the same subset of features are visible. Thus, the image of an object does not change qualitatively in terms of the presence of features over a *generic sector*. The boundaries of the generic sectors are places where small changes in the viewing direction produce abrupt changes in the features and the views at these points are said to be *singular views*.

A *generic view* of an object is a representation of the features which are visible for a *generic sector*. The representation for a *generic view* (as opposed to a 3D surface or volumetric representation) can be as simple as a bit map indicating which features are visible; or it can be more complex such as a hierarchical graph structure which makes explicit the relations between the visible features at different scales and includes analytic formulae for recovering object orientation from measurements of these features and relations between them. In the current implementation by Burns [BUR88], [BUR87a], [BUR87b] for prisms, a graph structure is used in which the nodes are lines and the arcs are binary relations between lines. Thus, both the nodes and arcs of the graphs are the features of a view.

For perspective projection, three-dimensional space is divided up into volumes. In practice, one has upper and lower bounds on the distance from the camera to the object, so that the space to be divided is actually finite. These bounds can come from several sources; one may have *a priori* knowledge of the environment, the object features themselves may not be resolvable by the camera at certain distances, and since there are only a finite number of transitions which can occur due to changes in distance, one can compute where they occur and find the largest [STE87].

One of the most important issues in object recognition is how to organize large model bases of complex objects so that the representation does not become unwieldy. This

becomes evident for the generic view approach since the complexity of the view sphere for n boolean features is $O(n^2)$ even for convex polyhedra [DYE87], [MAL88], [STE87]. One way in which this complexity is avoided is by computing the view spheres for single features and combining only the most stable and distinctive features [HEN87]. Thus, it is not necessary to compute the complete view sphere using every feature for each object.

For any recognition problem there will be a set of image features which is the basis for indexing and for recognition. These features could be points (e.g. vertices), lines, or regions; however, structures composed of these primitive tokens in particular relationships are even more powerful. For polyhedral objects these geometric structures might be parallel lines, parallelograms, regions bounded by straight lines, or any aggregation of tokens satisfying specific relations.

At UMass we are developing a recognition system based on binary relations between lines. Currently, models which are specified as planar surfaces, edges, and vertices, are compiled via generic views into a graph structure called a *prediction hierarchy*. A *prediction* is a statement or predicate concerning features of the image of an object. For example, it may be as simple and general as an assertion that a pair of line segments in the projection are parallel; or as complex as constraints on the relative orientations and distances between all pairs of line segments which are simultaneously visible. A prediction is represented here as a relational graph; in the initial implementation, the elements in the graph are projected straight-line segments. The relations associated with arcs in the graph are constraints on the relative orientations, positions and lengths of a pair of segments. The constraints define an extent box in the four-dimensional parameter space. The line-segment relations described by extent boxes form the primitive features which are combined to form the prediction hierarchy.

The features which are used in a prediction hierarchy are selected automatically based on an analysis of the viewing spheres for all of the objects in the model base. These features are ranked based on two factors. One of those factors is the size of the extent box; the other is the visibility. A feature is considered useful if its extent box is small in volume and it is visible over a wide range in viewpoints (for example, the two view-invariant relations: parallel and endpoint coincidence). Having a small extent box is important

if the relation is to help characterize an object's projection with a specificity sufficient to discriminate the object from a large number of other objects and from chance arrangements of image segments. Although invariant relations are clearly useful [LOW85] they alone are not sufficient to fully characterize projections. For instance, proportions are often strong characterizations of object structure, but the length measurement ratios that represent them are often not strictly view invariant. For example, a tall box has a height to width ratio that is significantly different from a cube over a large range in views.

It should be clear from the above discussion that a prediction may be valid only over a restricted set of views for a given object. A prediction *instance* is a set of model segments, a mapping from the model segments to the segments of the prediction's relational graph and the range of viewpoints from which the prediction is valid for these segment bindings. For a given model base, each prediction has a set of such instances and a *cumulative visibility*, the total area of all their visibility regions on the viewing sphere, across all objects. The prediction hierarchy is intended to be an efficient representation of all of the views of all of the objects in the model base. In general, there will be many simple structures which are shared by different views of an object or by different objects. As a result, we expect significant savings in the number of structures which need to be represented [BUR88]. The prediction hierarchy is used in the matching process and it provides a natural structure for a flexible control strategy in the matching process.

REFERENCES

ADI85a G. Adiv, "Determining 3-D Motion and Structure from Optical Flow Generated by Several Moving Objects," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Volume PAMI-7, July 1985, pp. 384-401.

ADI85b G. Adiv, "Interpreting Optical Flow," Ph.D. Dissertation, Computer and Information Science Department, University of Massachusetts at Amherst, September 1985.

ADI85c G. Adiv, "Inherent Ambiguities in Recovering 3-D Motion and Structure from a Noisy Flow Field," *Proc. of the Computer Vision and Pattern Recognition Conference*, San Francisco, CA, June 1985, pp. 7077.

ANA84 P. Anandan, "Computing Dense Displacement Fields with Confidence Measures in Scenes Containing Occlusion," *SPIE Intell gen. Robots and Computer Vision Conference*, Volume 521, 1984, pp. 184-194; also *DARPA IU Workshop Proceedings*, 1984; and COINS Technical Report 84-32, University of Massachusetts at Amherst, December 1984.

ANA85a P. Anandan and R. Weiss, "Introducing a Smoothness Constraint in a Matching Approach for the Computation of Optical Flow Fields," *Proc. of the Third Workshop on Computer Vision: Representation and Control*, October 1985, pp. 186-196, also in *DARPA IU Workshop Proceedings*, 1985.

ANA85b P. Anandan, "Motion and Stereopsis," COINS Technica: Report 85-52, University of Massachusetts at Amherst, December 1985; also to appear (in Spanish) in *Vision por Computador*, (Carme Torras, ed.), to be published by Alianza Editorial, Spain.

ANA87a P. Anandan, "Measuring Visual Motion From Image Sequences", Ph.D. Dissertation, University of Massachusetts at Amherst, January 1987.

ANA87b P. Anandan, "A Unified Perspective on Computational Techniques for the Measurement of Visual Motion", *Proc. of the International Conference on Computer Vision*, London, England, June 1987.

ANA88 P. Anandan, "A Computational Framework and an Algorithm for the Measurement of Visual Motion", to appear in *International Journal on Computer Vision*, 1988.

ARK87a R. Arkin, "Motor Schema-Based Navigation for a Mobile Robot: An Approach to Programming by Behavior", *Proc. of the IEEE International Conference on Robotics and Automation*, Raleigh, NC, pp. 264-271, March 1987.

ARK87b R. Arkin, "Towards Cosmopolitan Robots: Intelligent Navigation in Extended Man-Made Environments", Ph.D. Thesis, Computer & Information Science Department, University of Massachusetts at Amherst, September 1987.

ARK87c R. Arkin, A. Hanson, and E. Riseman, "Visual Strategies for Mobile Robot Navigation", *Proc. of IEEE Computer Society Workshop on Computer Vision*, Miami, FL November, 1987.

BAL87a P. Balasubramanyam, "Extraction of Motion In Depth: A First Step in Stereoscopic Motion Interpretation", presented at the American Institute of Aeronautics and Astronautics Computers in Aerospace IV Conference, Wakefield, MA, October 1987, pp. 277-286.

BAL87b P. Balasubramanyam, "Computation of Motion-In-Depth Parameters Using Stereoscopic Motion Constraints", *Proc. of IEEE Computer Society Workshop on Vision*, Miami, FL, November 1987.

BAL88 P. Balasubramanyam and M. Snyder, "Computation of Motion in Depth Parameters: A First Step in Stereoscopic Motion Interpretation", *Proc. of the DARPA Image Understanding Workshop*, Cambridge, MA, April 1988.

BEL86 R. Belknap, E. Riseman, and A. Hanson, "The Information Fusion Problem and Rule-Based Hypotheses Applied to Complex Aggregations of Image Events", *Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Miami, FL, June 1986, pp. 227-234.

BEV87 J. Ross Beveridge, J. Griffith, R. Kohler, A. Hanson, and E. Riseman, "Segmenting Images Using Localized Histograms and Region Merging", COINS Technical Report 87-88, University of Massachusetts at Amherst, September 1987.

BHA85 S. Bharwani, A. Hanson, and E. Riseman, "Refinement of Environmental Depth Maps over Multiple Frames," *Proc. DARPA IU Workshop*, Miami Beach, FL, December 1985.

BHA86 S. Bharwani, E. Riseman, and A. Hanson, "Refinement of Environmental Depth Maps Over Multiple Frames", *Proc. of the Workshop on Motion: Representation and Analysis*, Charleston, SC, May 1986, pp. 73-80.

BOL87 M. Boldt and R. Weiss, "Token-Based Extraction of Straight Lines", COINS Technical Report 87-104, University of Massachusetts at Amherst, October 1987.

BRO88 J. Brolio, J. Ross Beveridge, B. Draper, J. Griffith, G. Reynolds, R. Collins, J. Burrill, L. Williams, A. Hanson, E. Riseman, "Symbolic Data Management

for Vision Research", COINS Technical Report, University of Massachusetts at Amherst, 1988.

[BUR86] J.B. Burns, A.R. Hanson, and E.M. Riseman, "Extracting Straight Lines," *IEEE Transactions on Pattern Analysis and Machine Intelligence 8*, No. 4, July 1986, 425-455.

[BUR87a] J.B. Burns, "Recognition in 2D Images of 3D Objects from Large Model Bases Using Prediction Hierarchies", *Proc. of the International Joint Conference on Artificial Intelligence*, Milan, Italy, August 1987, pp. 763-766.

[BUR87b] J.B. Burns and L.J. Kitchen, "An Approach to Recognition in 2D Images of 3D Objects From Large Model Bases", COINS Technical Report 87-85, University of Massachusetts at Amherst, August 1987.

[BUR88] J.B. Burns and Leslie J. Kitchen, "Rapid Object Recognition From A Large Model-Base Using Prediction Hierarchies", *Proc. DARPA Image Understanding Workshop*, Cambridge, MA, April 1988. Also COINS Technical Report, University of Massachusetts at Amherst, in preparation.

[CAL85] J. Callahan and R. Weiss, "A Model For Describing Surface Shapes", *CVPR85*, pp. 240-245, 1985.

[CHA82] I. Chakravarty and H. Freeman, "Characteristic Views as a Basis For Three-Dimensional Object Recognition", *SPIE*, pp. 37-45, 1982.

[DEM68] A.P. Dempster, "A Generalization of Bayesian Inference", *Journal of the Royal Statistical Society*, Series B., Vol. 30, 1968, pp. 205-247.

[DOL86] J. Dolan, G. Reynolds, and L. Kitchen, "Piecewise Circular Description of Image Curves Using Constancy of Grey-Level Curvature", COINS Technical Report 86-33, University of Massachusetts at Amherst, July 1986.

[DOL88] John Dolan and Richard Weiss, "The Perceptual Organization of Image Curves", COINS Technical Report, University of Massachusetts at Amherst, in preparation.

[DRA87a] B. Draper, R. Collins, J. Brolio, J. Griffith, A. Hanson, and E. Riseman, "Tools and Experiments in the Knowledge-Directed Interpretation of Road Scenes", *Proc. of the DARPA Image Understanding Workshop*, Los Angeles, CA, February 1987, pp. 178-193.

[DRA87b] B. Draper, R. Collins, J. Brolio, A. Hanson, and E. Riseman, "The Schema System", COINS Technical Report, University of Massachusetts at Amherst, in preparation.

[DRA88] B. Draper, J. Brolio, R. Collins, A. Hanson, and E. Riseman, "Information Fusion by Distributed Subsystems in a Knowledge-Based Vision Architecture", to appear in *Computer Vision and Pattern Recognition*, 1988.

[DUT88] R. Dutta, R. Manmatha, E. Riseman, and M.A. Snyder, "Issues in Extracting Motion Parameters and Depth from Approximate Translational Motion", *Proc. DARPA Image Understanding Workshop*, Cambridge, MA, April 1988.

[DYE87] H. Plantinga and C. Dyer, "Visibility, Occlusion, and the Aspect Graph", Report 736, University of Wisconsin–Madison, December 1987.

[FEK84] G. Fekete and L.S. Davis, "Property Spheres: A New Representation for 3D Object Recognition", *CVW84*, pp. 192-204, 1984.

[FIS86] M.A. Fischler and R.C. Bolles, "Perceptual Organization and Curve Partitioning", *IEEE PAMI*, 8(1), pp. 100-105, January 1986.

[GLA87a] F. Glazer, "Hierarchical Motion Detection", Ph.D. Dissertation, Computer and Information Science Department, University of Massachusetts, Amherst, February 1987.

[GLA87b] F. Glazer, "Hierarchical Gradient-Based Motion Detection", COINS Technical Report 87-77, University of Massachusetts at Amherst, 1987.

[GLA87c] F. Glazer, "Computation of Optic Flow by Multilevel Relaxation", Coins Technical Report 87-64, University of Massachusetts at Amherst, 1987.

[HAN78a] A.R. Hanson and E.M. Riseman (Eds.), *Computer Vision Systems*, New York, Academic Press, 1978.

[HAN78b] A. Hanson and E. Riseman, "Segmentation of Natural Scenes," in *Computer Vision Systems*, (A. Hanson and E. Riseman, Eds.), Academic Press 1978, pp. 129-163.

[HAN78c] A. Hanson, and E. Riseman, "VISIONS: A Computer System for Interpreting Scenes, in *Computer Vision Systems* (A. Hanson and E. Riseman, Eds.) pp. 303-333, Academic Press, 1978.

[HAN86] A.R. Hanson and E.M. Riseman, "A Methodology for the Development of General Knowledge-Based Vision Systems," in *Vision, Brain, and Cooperative Computation*, (M. Arbib and A. Hanson, Eds.) 1986, MIT Press, Cambridge, MA.

[HAN87a] A.R. Hanson and E.M. Riseman, "The VISIONS Image Understanding System", in *Advances in Computer Vision*, Chris Brown, (Ed.), Erlbaum Press, 1987. Also COINS Technical Report 86-62, University of Massachusetts at Amherst, December 1986.

[HAN87b] A. Hanson and E. Riseman, "From Image Measurements to Object Hypotheses", COINS Technical Report 87-129, University of Massachusetts at Amherst, December 1987.

[HEN87] C. Hanson and T. Henderson, "CAGD-Based Computer Vision", *IEEE Workshop on Computer Vision*, Miami Beach, FL, November 1987, pp. 100-105.

[IKE87] K. Ikeuchi, "Precompiling A Geometrical Model Into An Interpretation For Object Recognition in Bin-Picking Tasks:, *IUW87*, pp. 321-330, 1987.

[KAH87] P. Kahn, L. Kitchen, and E. Riseman, "Real-Time Feature Extraction: a Fast Line Finder for Vision-Guided-Robot Navigation", COINS Technical Report 87-57, University of Massachusetts at Amherst, July 1987.

[KER81] Y.L. Kergosien, "La Famille des Projections Orthogonales, d'une Surface et ses Singularities", *Comptes Randus Acad. Sc. Paris*, 292, pp. 929-932, 1981.

[KOE76] J.J. Koenderink and A.J. van Doors, "The Singularities of the Visual Mapping", *BC*, Vol., 24, No. 1, pp. 51-59, 1976.

[KOE79] J.J. Koenderink and A.J. van Doors, "Internal Representation of Shape with Respect to Vision", *BV*, Vol. 32, No. 4, pp. 211-216, 1979.

[KOE84] J.J. Koenderink, "What Does the Occluding Contour Tell Us About Solid Shape?", *Perception*, Vol. 13, pp. 321-330, 1984.

[KOH87] C.A. Kohl, A.R. Hanson, and E.M. Riseman, "A Goal-Directed Intermediate Level Executive for Image Interpretation", *Proc. of the International Joint Conference on Artificial Intelligence*, Milan, Italy, August 1987, pp. 811-814.

[KOH88] C.A. Kohl, "Goal-Directed Control for Computer Vision", Ph.D. Dissertation, Computer & Information Science, University of Massachusetts at Amherst, February 1988.

[LAW83] D.T. Lawton, "Processing Translational Motion Sequences", *Computer Graphics and Image Processing*, Vol. 22, pp. 116-144, 1983.

[LAW84] D.T. Lawton, "Processing Dynamic Image Sequences from a Moving Sensor," Ph.D. Dissertation (COINS Technical Report 84-05), Computer and Information Science Department, University of Massachusetts, 1984.

[LEH87] N. Lehrer, G. Reynolds, and J. Griffith, "A Method for Initial Hypothesis Formation in Image Understanding", COINS Technical Report, University of Massachusetts at Amherst, in preparation, 1987.

[LOW85] D.G. Lowe, *Perceptual Organization and Visual Recognition*, Kluwer Academic Publishers, 1985.

[MAL88] Z. Gigus, and J. Malik, "Computing the Aspect Graph for Line Drawings of Polyhedral Objects," Report UCB-CSD 88-402, University of California, Berkeley, February 1988.

[PAV85] I. Pavlin, A. Hanson, and E. Riseman, "Analysis of an Algorithm for Detection of Translational Motion," *Proc. DARPA IU Workshop*, Miami Beach, FL, December 1985.

[PAV86] I. Pavlin, E. Riseman, and A. Hanson, "Translational Motion Algorithm With Global Feature Constraints", COINS Technical Report 86-58, University of Massachusetts at Amherst, December 1986.

[PAV88] I. Pavlin, "Motion From a Sequence of Images", *Proc. of the DARPA Image Understanding Workshop*, Cambridge, MA, April 1988.

[REY84] G. Reynolds, N. Irwin, A. Hanson and E. Riseman, "Hierarchical Knowledge-Directed Object Extraction Using a Combined Region and Line Representation," *Proc. of the Workshop on Computer Vision: Representation and Control*, Annapolis, Maryland, April 30 - May 2, 1984, pp. 238-247.

[REY85] G. Reynolds, D. Strahman, N. Lehrer, "Converting Feature Values to Evidence," *Proc. DARPA IU Workshop*, Miami Beach, FL, 1985.

[REY86a] G. Reynolds, D. Strahman, N. Lehrer, L. Kitchen, "Plausible Reasoning and the Theory of Evidence", COINS Technical Report 86-11, University of Massachusetts at Amherst, April 1986.

[REY86b] G. Reynolds, J. Ross Beveridge, "Geometric Line Organization Using Spatial Relations and a Connected Components Algorithm", COINS Technical Report, University of Massachusetts at Amherst, in preparation, 1986.

[REY87] G. Reynolds and J.R. Beveridge, "Searching For Geometric Structure in Images of Natural Scenes", COINS Technical Report 87-03, University of Massachusetts at Amherst, January 1987.

[RIS84] E.M. Riseman and A.R. Hanson, "A Methodology for the Development of General Knowledge-Based Vision Systems," *IEEE Proc. of the Workshop on Computer Vision: Representation and Control*, 1984, pp. 159-170.

[RIS87] E. Riseman, A. Hanson, and R. Belknap, "The Information Fusion Problem: Forming Token Aggregations Across Multiple Representations", COINS Technical Report 87-48, University of Massachusetts at Amherst, 1987.

[ROS87] A. Rosenfeld, "A Report on the DARPA Image Understanding Architectures Workshop", *Proc. of the DARPA Image Understanding Workshop*, Los Angeles, CA, February 1987, pp. 298-302.

[SHA76] G. Shafer, "A Mathematical Theory of Evidence", Princeton University Press, 1976.

[SHA86] H. Shariat, "The Motion Problem: How To Use More Than Two Frames", Ph.D. Dissertation, University of Southern California, Los Angeles. Also Technical Report IRIS 202, Institute for Robotics and Intelligent Systems, University of Southern California, Los Angeles, CA, October 1986.

[SNY86] M. A. Snyder, "The Accuracy of 3D Parameters in Correspondence Based Techniques," *Proc. of the Workshop on Motion: Representation and Analysis*, Charleston, S.C., May 1986; also COINS Technical Report 86-28, University of Massachusetts at Amherst, July 1986.

[SNY88] M.A. Snyder, "The Precision of 3D Parameters in Correspondence-Based Techniques: The Case of Uniform Translational Motion in a Rigid Environment", to appear in *IEEE PAMI*, 1988.

[STE87] J. Stewman and K. Bowyer, "Aspect Graphs For Convex Planar-Face Objects", *Proc. IEEE Workshop on Computer Vision: Representation and Control*, Miami Beach, pp. 123-130, November 1987.

[THOR87] C. Thorpe, S. Shafer, and T Kanade, "Vision and Navigation for the Carnegie Mellon Navlab", *Proc. of the DARPA Image Understanding Workshop*, Los Angeles, CA, February 1987, pp. 143-152.

[THOM87] D. Thompson and J. Mundy, "Three Dimensional Model Matching From an Unconstrained Viewpoint, *Proc. of the IEEE Conference on Robotics and Automation*, Raleigh, NC, 1987.

[ULL79] S. Ullman, "The Interpretation of Visual Motion", MIT Press, Cambridge, MA, 1979.

[WAX86] A.M. Waxman and J.H. Duncan, "Binocular Image Flows: Steps Toward Stereo-Motion Fusion", *IEEE PAMI*, Vol. PAMI-8, November 1986, pp. 715-729.

[WEE84] C. Weems, "Image Processing on a Content Addressable Array Parallel Processor", Ph.D. Dissertation and COINS Technical Report 84-14, University of Massachusetts at Amherst, September 1984.

[WEE87] C. Weems, S. Levitan, A. Hanson, and E. Riseman, "The Image Understanding Architecture Workshop", *Proc. of the DARPA Image Understanding Workshop*, Los Angeles, CA, February 1987, pp. 483-498.

[WEE88a] C. Weems, S. Levitan, A. Hanson. E. Riseman, G. Nash, and D. Shu, "The Image Understanding Architecture", to appear in The International Journal on Computer Vision, 1988. Also Technical Report 87-76, Computer & Information Science, University of Massachusetts at Amherst, August 1987.

[WEE88b] Charles Weems, "Some Sample Algorithms for the Image Understanding Architecture", *Proc. DARPA Image Understanding Workshop*, Cambridge, MA, April 1988.

[WEE88c] C. Weems, E. Riseman, A. Hanson, and A. Rosenfeld, "An Integrated Image Understanding Benchmark: Recognition of a 2 1/2 D 'Mobile'", *Proc. DARPA Image Understanding Workshop*, Cambridge, MA, April 1988.

[WEI85] R. Weiss, A. Hanson, and E. Riseman, "Geometric Grouping of Straight Lines," *Proc. 1985, DARPA IU Workshop,* Miami Beach, FL, 1985

[WEI86] R. Weiss and M. Boldt, "Geometric Grouping Applied to Straight Lines", *Proceedings on the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Miami, Florida, June, 1986, pp. 489-495.

[WES82] L. Wesley and A. Hanson, "The Use of Evidential-Based Model for Representing Knowledge and Reasoning About Images in the VISIONS System", *Proc. Workshop on Computer Vision*, Rindge, NH, August 1982.

[WES83] L. Wesley, "Reasoning About Control: The Investigation of an Evidential Approach", *Proc. 8th IJCAI*, Karlsruhe, West Germany, August 1983, pp. 203-210.

[WES86] L. Wesley, "Evidential Knowledge-Based Computer Vision", *Optical Engineering*, Vol. 25, No. 3, March 1986, pp. 363-379.

[WES88] L. Wesley, "Evidential-Based Control in Knowledge-Based Systems", Ph.D. Dissertation, Computer & Information Science, University of Massachusetts at Amherst, February 1988.

WEY86 T.E. Weymouth, "Using Object Descriptions in a Schema Network For Machine Vision," Ph.D. Dissertation, Computer and Information Science Department, University of Massachusetts at Amherst. Also COINS Technical Report 86-24, University of Massachusetts at Amherst, 1986.

[WIL86] Lance R. Williams and P. Anandan, "A Coarse-to-Fine Control Strategy for Stereo and Motion on a Mesh-Connected Computer", COINS Technical Report 86-19. University of Massachusetts at Amherst, May 1986.

[WIL88a] L. Williams and A. Hanson, "Depth From Looming Structure", *Proc. of the DARPA Image Understanding Workshop*, Cambridge, MA, April 1988.

[WIL88b] L. Williams and A. Hanson, "Translating Optical Flow Into Token Matches", *Proc. of the DARPA Image Understanding Workshop*, Cambridge, MA, April 1988.

[WIT83] A.P. Witkin and J.M. Tenenbaum, "What Is Perceptual Organization For?", *IJCAI*, 1983, pp. 1023-1026.